1. **R -SQUARED AND RESIDUAL SUM OF SQUARES(RSS)**
   R-squared measures the goodness of fit of a regression model. Because we feel that large negative residuals (i.e., points far below the line) are as bad as large positive ones (i.e., points that are high above the line). By squaring the residual values, we treat positive and negative discrepancies in the same way.
2. **TSS, ESS AND RSS**
   The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model. The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.
   the **explained sum of squares** (ESS), alternatively known as the model sum of squares or sum of squares due to regression (SSR – not to be confused with the residual sum of squares (RSS) or sum of squares of errors), is a quantity used in describing how well a model, often a regression model, represents the data being modelled.
   The Sum of Squared regression is the sum of the differences between the predicted value and the mean of the dependent variable. The Sum of Squared Error is the difference between the observed value and the predicted value
   TSS = ESS + RSS
3. **NEED OF REGULARIZATION IN MACHINE LEARNING**
   We need regularization in machine learning because regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.
4. **GINI IMPURITY INDEX**
   The Gini Index is a summary measure of income inequality. The Gini coefficient incorporates the detailed shares data into a single statistic, which summarizes the dispersion of income across the entire income distribution
   The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one. It favours mostly the larger partitions and are very simple to implement. In simple terms, it calculates the probability of a certain randomly selected feature that was classified incorrectly.
5. **UNREGULARIZED DECISION TREES**
   Yes, unregularized decision trees are prone to overfitting. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.
6. **ENSEMBLE TECHNIQUES**
   Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.
   Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.
7. **BAGGING AND BOOSTING TECHNIQUES**
   Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification. It attempts to increase the weight of an observation if it was erroneously categorized. Boosting creates good predictive models in general.
8. **OUT OF BAG ERROR**
   The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the Random Forest Classifier to be fit and validated whilst being trained

9. **K FOLD CROSS VALIDATION**
K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into.
It is a method that is easy to comprehend, works well for a limited data sample and also offers an evaluation that is less biased, making it a popular choice. The data sample is split into 'k' number of smaller samples, hence the name: K-fold Cross Validation.

10. **HYPER PARAMETER TUNING**
Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.
Hyperparameter tuning takes advantage of the processing infrastructure of Google Cloud to test different hyperparameter configurations when training your model. It can give you optimized values for hyperparameters, which maximizes your model's predictive accuracy.

11. **GRADIENT DESCENT – LEARNING RATE**
In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.

12. **LOGISTIC REGRESSION**
We can’t use logistic regression for classification of nonlinear data. The reason is that the target label has no linear correlation with the features. In such cases, logistic regression (or linear regression for regression problems) can't predict targets with good accuracy (even on the training data).

13. **ADABOOST AND GRADIENT BOOST**
AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.

14. **BIAS VARIANCE TRADE OFF**
In statistics and machine learning, the bias–variance trade-off is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

15. **LINEAR KERNEL**
Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set.

RBF KERNEL
In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

POLYNOMIAL KERNEL
In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models